



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Prediction of the Water Content in Protein Binding Sites

**Citation for published version:**

Michel, J, Tirado-Rives, J & Jorgensen, WL 2009, 'Prediction of the Water Content in Protein Binding Sites', *Journal of Physical Chemistry B (Soft Condensed Matter and Biophysical Chemistry)*, vol. 113, no. 40, pp. 13337-13346. <https://doi.org/10.1021/jp9047456>

**Digital Object Identifier (DOI):**

[10.1021/jp9047456](https://doi.org/10.1021/jp9047456)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of Physical Chemistry B (Soft Condensed Matter and Biophysical Chemistry)

**Publisher Rights Statement:**

Copyright © 2009, American Chemical Society

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Published in final edited form as:

*J Phys Chem B*. 2009 October 8; 113(40): 13337–13346. doi:10.1021/jp9047456.

## Prediction of the Water Content in Protein Binding Sites

Julien Michel, Julian Tirado-Rives, and William L. Jorgensen\*

Department of Chemistry, Yale University, New Haven CT-06520, USA

### Abstract

An efficient molecular simulation methodology has been developed to determine the positioning of water molecules in the binding site of a protein or protein-ligand complex. Occupancies and absolute binding free energies of water molecules are computed using a statistical thermodynamics approach. The methodology, referred to as JAWS, features “ $\theta$ -water” molecules that can appear and disappear on a binding-site grid. Key approximations render the technique far more efficient than conventional free energy simulations. Testing of JAWS on five diverse examples (neuraminidase, scytalone dehydratase, major urinary protein 1,  $\beta$ -lactoglobulin, and COX-2) demonstrates its accuracy in locating hydration sites in comparison to results from high-resolution crystal structures. Possible applications include aid in refinement of protein crystal structures, drug lead optimization, setup of docking calculations and simulations of protein-ligand complexes.

### Keywords

water in proteins; protein-ligand interactions; protein hydration; free energy simulations

### Introduction

The accurate computation of free energies of binding is a key objective of computer-aided drug design.<sup>1,2</sup> A potent ligand usually shows good steric and electrostatic complementary with its receptor, but other factors such as conformational or vibrational entropy,<sup>3,4</sup> ligand strain,<sup>5</sup> and desolvation,<sup>6</sup> play significant roles. The important contribution of the aqueous medium to ligand binding is also apparent from the hydrophobic effect and numerous observations of water mediated protein-ligand interactions in binding sites.<sup>7,8</sup> Indeed, the displacement of water molecules in binding sites by judicious ligand modification has emerged as a strategy to optimize lead compounds.<sup>9,10</sup>

Implicit solvent theories can be used to model successfully the effect of bulk desolvation in protein-ligand binding,<sup>11–13</sup> but they do not account for specific water-ligand interactions. Indeed, explicit consideration of a few water molecules in the vicinity of a ligand has been shown to improve the quality of predictions from docking algorithms.<sup>14–17</sup> Given that information about the locations of hydration sites is often missing in docking exercises, or uncertain due to difficulties in resolving water molecules by crystallography,<sup>18–20</sup> alternative computational methodologies have been sought.

Empirical techniques based on interactions energies or QSAR descriptors have been reported.<sup>21–24</sup> While rapid, these techniques are often of limited accuracy or suffer from lack of transferability. For instance, the method CMIP has been reported to overestimate the number

\*To whom correspondence should be addressed. William.Jorgensen@yale.edu.

Supporting Information Available: Implementation details for the clustering procedure. This information is available free of charge via the Internet at <http://pubs.acs.org>.

of hydration sites, possibly because it neglects entropic contributions to the affinity of a water molecule for a given site.<sup>22,25</sup> Alternatively, Monte Carlo (MC) or molecular dynamics (MD) simulation can be conducted to equilibrate the water distribution in a binding site. In particular, Lazaridis and Li have used MD simulations and inhomogeneous fluid solvation theory to evaluate the binding enthalpies and entropies for interfacial water molecules in protein-ligand complexes.<sup>26</sup> The approach has been significantly extended to both locate all water molecules in a protein binding site and evaluate the favorability of their displacement.<sup>27,28</sup> Again, such information is expected to be useful for ligand design and optimization.<sup>9</sup> A major problem with MC/MD approaches is that diffusion of water molecules in and out of cavities at the protein interface can be excessively slow. Depending on the nature of the binding site, it may be impossible using standard simulation methodologies to obtain results that do not depend markedly on the initial setup.

Grand canonical Monte Carlo methods overcome this problem by letting the number of water molecules fluctuate during a simulation.<sup>29</sup> The approach allows, in principle, water molecules to exchange in and out of cavities with bulk solvent. However, even with cavity bias and configurational bias moves,<sup>30,31</sup> insertions and deletions of entire water molecules are very infrequent (<1% acceptance rate).<sup>32</sup> Additionally, computation of a free energy change in the  $\mu$ VT ensemble does not correspond to the free energy changes under the usual experimental NPT conditions. Thus, more complex simulation protocols are needed to select a constant chemical potential  $\mu$  that yields an ensemble closely resembling the experimental conditions; this typically requires time consuming simulations of proteins embedded in large solvent boxes.<sup>33</sup> The other main alternative based on free-energy simulations relies on the computation of the absolute binding free energy of a water molecule at a putative location; both free-energy perturbation (FEP) and thermodynamic integration methods have been applied to this end for 20 years.<sup>34–36</sup> The main obstacle is the number of simulations needed to compute the absolute binding free energy of many water molecules. Though the approach is conceptually sound, it is challenging for routine applications.

Thus, it is well accepted that accurate prediction of hydration sites is important to improve accuracy in modeling protein-ligand interactions. The issue has arisen repeatedly in our own lead-optimization efforts that have been guided by free energy perturbation (FEP) calculations.<sup>37–44</sup> In the simulations, the presence or absence of water molecules in locations between the protein and ligand often seem uncertain. Given the limitations of current techniques, a novel molecular simulation methodology to rapidly and accurately determine the water content of binding sites was sought. The outcome of these efforts is reported here with application to three protein-ligand complexes and the ligand-binding sites of two apo proteins. Performance measures include comparisons between the predictions and the observed positions of water molecules in protein crystal structures.

## Methods

### Theory

As described below, our approach to determine the optimal placement of water molecules in a binding site features the equilibration of the distribution of water molecules on a grid. However, to begin, equilibrium hydration of a binding site depends on the difference between the free energy of a water molecule in bulk solvent and in the binding site (Figure 1A and eq 1).

$$\Delta G_b(\text{water}) = G_{\text{water}}^{\text{site}} - G_{\text{water}}^{\text{bulk}} = -kT \ln \frac{Q_{\text{water}}^{\text{site}}}{Q_{\text{water}}^{\text{bulk}}} \quad (1)$$

However, estimation of this free energy difference by computer simulations, using the Zwanzig equation for instance,<sup>45</sup> is not practical as the exchange of water molecules between bulk and the binding site can be very slow. Alternatively, this quantity can be estimated using a series of unphysical intermediate systems that allow the transfer of a water molecule from bulk to the protein binding site (Figure 1B–E).

With this “double decoupling approach”,<sup>34</sup> which compares the removal of a water molecule from the bulk and from the binding site,<sup>46</sup> the free energy of binding the water molecule  $\Delta G_b(\text{water})$  can be written as eq 2.

$$\begin{aligned}\Delta G_b(\text{water}) &= -kT \ln \left( \frac{Q_{\text{bulk}}^{\text{ideal}}}{Q_{\text{water}}^{\text{bulk}}} \right) + -kT \ln \left( \frac{Q_{\text{ideal}}^{\text{site,constr}}}{Q_{\text{ideal}}^{\text{bulk}}} \right) + -kT \ln \left( \frac{Q_{\text{water}}^{\text{site,constr}}}{Q_{\text{ideal}}^{\text{site,constr}}} \right) + -kT \ln \left( \frac{Q_{\text{water}}^{\text{site}}}{Q_{\text{water}}^{\text{site,constr}}} \right) \\ &= -\Delta G_{\text{hyd}}(\text{water}) + \Delta G_{\text{constr}}(\text{ideal, site}) + \Delta G_{\text{trans}}(\text{water, site}) + \Delta G_{\text{constr}}(\text{water, site})\end{aligned}\quad (2)$$

The first term  $-\Delta G_{\text{hyd}}(\text{water})$  corresponds to the free energy change for removing the intermolecular interactions of a water molecule in bulk (see Figure 1B). It is the negative of the excess hydration free energy of water.<sup>47</sup> In this study the value +6.4 kcal/mol was adopted as a consensus from the computed free energies of hydration of the TIP4P water model,<sup>36,48,49</sup> and experimental data.<sup>50</sup>

The second term of eq 2,  $\Delta G_{\text{constr}}(\text{ideal, site})$ , corresponds to the free energy of constraining the now ideal particle (i.e., non-interacting) to occupy a volume  $V^{\text{constr}}$  in a binding site instead of the volume  $V^0$  available to a water molecule in bulk solvent (Figure 1C). Since the two states differ only in translational entropy, this term is equal to the ratio of available volumes as shown in eq 3,<sup>51</sup> where an appropriate value of  $V^0$  is the inverse of the concentration  $C^0$  of bulk water, 55.55 mol/L,<sup>47</sup> and  $V^{\text{constr}}$  depends on the nature of the constraint.

$$\Delta G_{\text{constr}}(\text{ideal, site}) = -kT \ln \left( \frac{V^{\text{constr}}}{V^0} \right) \quad (3)$$

The third term in eq 2,  $\Delta G_{\text{trans}}(\text{water, site})$ , corresponds to the conversion of the localized ideal particle into a water molecule (Figure 1D). A variety of free-energy techniques can be employed to compute this term. For example, the Lennard-Jones (LJ) terms may be first turned on, followed by the atomic partial charges on the water molecule. It is important that the particle is constrained; otherwise when the intermolecular interactions are removed, the particle could sample the entire simulation volume, which causes large numerical errors.<sup>52</sup> A hard-wall potential may be used whereby the particle is constrained to occupy a sphere of radius  $R$ , which must be judiciously chosen. If  $R$  is too large, the process will not be well reversible and numerical errors will arise. If  $R$  is too small, important configurations of the fully interacting water molecule will be missed.<sup>34,53</sup> For the calculations reported here,  $R = 2.8 \text{ \AA}$  was found to be adequate.

The last term in eq 2,  $\Delta G_{\text{constr}}(\text{water, site})$ , corresponds to the free energy change for removing the constraint (Figure 1E). However, without such a term, the location of the water in the binding site is ill defined, i.e., it could exchange with other water molecules or drift away from a weak affinity site, which would cloud the interpretation of the computed binding affinities. Since binding free energies to regions with the same, small volumes are sought here, it is appropriate to consider them as ‘local binding free energies’ and  $\Delta G_{\text{constr}}(\text{water, site}) = 0 \text{ kcal/mol}$  in the present analysis.

The analysis above can be extended to simulate the insertion of water molecules into N hydration sites (with indices  $i = 1, \dots, N$ ) in a binding site via eq 4. Evaluation of eq 4 is complicated by the

$$\Delta G_b(N \text{ waters}) = -kT \ln \left( \frac{Q_{sites, constr}^{N \text{ waters}}}{\prod_{i=1}^N Q_{ideal}^{site i, constr}} \right) + \left( \sum_{i=1}^N \Delta G_{constr}(ideal, site i) - \Delta G_{hyd}(water) \right) \quad (4)$$

existence of coupled interactions between water molecules located at each hydration site. In principle, N water molecules could be iteratively transferred to an increasing number of hydration sites ( $i = 1, 2, \dots, N$ ) until a global minimum in  $\Delta G_b$  is found at which stage the optimum integer number of water molecules has been determined. This approach is time-consuming owing to the large number of required free-energy calculations. Furthermore, prior knowledge of the positions of each of the N putative hydration sites in the binding site would be needed, and alternative distributions of the same number of hydration sites have to be considered.

Fortunately, approximations can be elaborated from eq 4 to yield a more practical methodology. The usual expression for the potential energy function  $U(r)$  that describes N water molecules in a protein binding site can be modulated using N scaling parameters  $\theta_i$ , as shown in eq 5, where  $U_{inter}$  is the intermolecular energy of water molecule i and  $U_0$  gives the remaining energy terms. In the present

$$U(r, \sum_{i=1}^N \theta_i) = U_0(r) + \sum_{i=1}^N \theta_i U_{inter}(r, water i) \quad (5)$$

approach, the  $\theta_i$  are treated as degrees of freedom, which can be sampled during a MC simulation in the same spirit as in the  $\lambda$ -dynamics method developed by Kong and Brooks.<sup>54</sup> The symbol  $\theta_i$  is used to distinguish this set of parameters from the coupling parameter  $\lambda$  used in FEP studies. The “ $\theta$ -water” i behaves as an ideal particle when  $\theta_i = 0$  and as a regular water molecule when  $\theta_i = 1$ . By collecting statistics during an MC simulation, the probability that a  $\theta$ -water is “water like” or “ideal like” is readily determined. The ratio of these two probabilities is formally related to the free energy change  $\Delta G_{trans}(water, site i)$  for transferring a water molecule from the gas phase into site i using eq 6.

$$\Delta G_{trans}(water, site i) = -kT \ln \left( \frac{P(\theta_i \rightarrow 1)}{P(\theta_i \rightarrow 0)} \right) \quad (6)$$

However, reliable free-energy estimates can only be obtained with eq 6 if the  $\theta_i$ -water molecule samples readily both high and low  $\theta$  values over the course of the simulation. If high energy barriers must be crossed, or if the free energy difference is large, excessive computational resources might be required before a statistically significant number of transitions could be observed. This difficulty can be overcome by adding a biasing term,  $V(\theta_i)$  in eq 7, for each of the N  $\theta$ -water molecules to the potential energy in eq 5.<sup>55</sup>

$$V(\theta_i) = \left( -\Delta G_{hyd} + \Delta G_{constr}(ideal, site i) \right) \theta_i \quad (7)$$

The  $V$  terms correspond exactly to what is necessary to correct eq 6 and estimate a binding free energy for each of the  $\theta$ -water molecules. In other words, addition of the  $V$  terms and collecting statistics about the probabilities that a  $\theta_i$  parameter is unity or zero, permits the direct partitioning of  $N$   $\theta$ -water molecules between bulk and the binding site. Each  $V(\theta_i)$  term penalizes high ‘water-like’  $\theta_i$  values by an amount that accounts for the free energy change for desolvation and localization of a water molecule at hydration site  $i$  (Figures 1B, 1C and first and second terms of eq 2). Therefore, evaluation of the ratio of probabilities of high and low  $\theta_i$  values (as in eq 6) during a simulation with the  $V(\theta_i)$  terms activated allows direct estimation of a free energy of binding  $\Delta G_b$  (as in eq 2) for each  $\theta$ -water, in the presence of  $N-1$  other  $\theta$ -water molecules.

The incorporation of the  $V$  biasing terms into the potential energy function before monitoring the  $\theta_i$  values to obtain a free energy of binding directly is more advantageous than attempting to correct post simulation the transfer free energy from eq 6. Even if a water molecule at a possible hydration site is never turned off or on completely during the simulation and hence its binding free energy is undetermined, the  $\theta_i$  provides a good indication whether this site would be occupied or unoccupied by a water molecule. By contrast, the correction terms could not be added post simulation, if insufficient statistics were collected to compute a transfer free energy using eq 6. Additionally, eq 7 provides the benefits of a biasing potential by penalizing high values of  $\theta$  which correspond to “water-like” states. Eq 7 therefore facilitates transitions between high and low  $\theta$  values and enhances convergence of the binding affinities.

The use of  $\theta$  moves to convert particles from “water-like” to “ideal-like” states has important benefits over the more typical FEP approach. As several water molecules can be assigned a  $\theta$  coupling parameter, a single simulation is sufficient to determine multiple hydration sites. In addition, cooperative hydration of the binding site by clusters of water molecules naturally arises over the course of the simulation. Notably, estimates of  $\Delta G_b$  for each water molecule can also be obtained rapidly from a single simulation, while decoupling dozens of water molecules serially by FEP would require substantially more computer time. Unlike grand-canonical Monte Carlo simulations, sampling of the extreme values of  $\theta$  is efficient as a change from a low to a high  $\theta$  value can happen gradually over several attempted moves. A potential issue with this approach is whether the presence of intermediates with partial “water-like” or “ideal-like” characteristics could affect the accuracy of the results. This is addressed here by validation studies.

## Implementation

At the outset, the binding site must be defined. Typically, it is for an apo protein or a protein-ligand complex. The present approach is to use a three-dimensional grid, which is formed by overlapping spheres centered on each atom including hydrogens of the ligand or on an alternative, user-selected set of atoms in the protein or complex. The radius of the spheres is normally 2–5 Å with a default of 3 Å. The resulting volume is filled with grid points spaced regularly in three dimensions; a 1 Å spacing is applied by default. The algorithm then consists of two distinct phases: finding the potential hydration sites, and determining their occupancies. The full system that is simulated consists of the protein, any bound ligand,  $\theta$ -water molecules, and regular water molecules.

In the first phase, putative hydration sites are detected by placing  $N$   $\theta$ -water molecules randomly on the grid. The number  $N$  can be manually specified; otherwise a default of 1 water molecule per 30 Å<sup>3</sup> of grid volume is added randomly onto the grid. This corresponds to the density of liquid water at 298 K and 1 atm, a reasonable upper estimate of the water density in the binding site. An MC simulation is performed that includes allowing the  $N$   $\theta$ -water molecules to sample freely the entire grid without any biasing potentials  $V(\theta_i)$ . An attempted MC move of a  $\theta$ -water includes rigid body translations and rotations and, for 50% of the



attempted moves, a random variation of its  $\theta_i$  value. In order to determine the most likely hydration sites, the  $\theta_i$  values are inspected every 100 MC steps; if a  $\theta_i$  is greater than a threshold value (0.95 by default), the  $\theta_i$ -water molecule is considered “water-like” and a counter on the nearest grid point is incremented by 1. As the simulation proceeds, the statistics collected in this fashion on the grid are interpreted as a probability distribution of water occupancies. At the end of the simulation, the fractional water occupancies are converted into an integer number of potential hydration sites using a clustering procedure (see Supporting Information).

In the second phase of the algorithm,  $\theta$ -water molecules are positioned at each of the potential hydration sites and constrained to occupy a volume defined by the nearest ( $\pm 1, \pm 1, \pm 1$ ) grid points, which corresponds to a  $27\text{-}\text{\AA}^3$  cube with a spacing of 1  $\text{\AA}$ . The biasing potentials  $V(\theta_i)$  are turned on and statistics about the  $\theta$  values are collected in a new MC simulation. A binding free energy for each hydration site is then estimated from the ratio of probabilities of observing a  $\theta$ -water at high ( $\theta > 0.95$ ) or low ( $\theta < 0.05$ )  $\theta$  values. Finally, at the end of the simulation, real water molecules are positioned at each hydration site with negative binding free energy. The orientation of these water molecules is optimized by means of a short MC simulation of  $100 \times N$  configurations at low temperature in which only the position and orientation of the water molecules are sampled. A coordinate file with the resultant well-hydrated complex is then generated for use in subsequent simulations. The methodology has been named Just Add Water Molecules or JAWS and has been implemented in a modified version of MCPRO, v. 2.1.<sup>56</sup>

Validation has been carried out by applying the methodology to a variety of protein-ligand complexes and apo proteins for which there are available high-resolution crystal structures. Standard procedures have been followed for the MC simulations including use of the OPLS-AA force field for the proteins and OPLS/CM1A for the ligands.<sup>57</sup> The CM1A atomic charges are scaled by 1.14 for neutral ligands and unscaled for charged ligands.<sup>58</sup> The Monte Carlo simulations are performed using Metropolis sampling.<sup>59</sup> Initial coordinates for the protein-ligand complexes are obtained from the Protein Data Bank. The ca. 200–300 protein residues with an atom within ca. 15  $\text{\AA}$  of a ligand atom are retained. Depending on the size of the binding site, 30–60  $\theta$ -water molecules are placed on the binding-site grid, and ca. 500–750 TIP4P water molecules are included in a ca. 22- $\text{\AA}$  radius cap.

The first stage of JAWS begins with an initial equilibration of the water molecules using 5 million MC configurations in which only water molecules are allowed to move. Then, the run to locate the hydration sites is performed using 5–15 million MC configurations with sampling of all water molecules, the  $\theta_i$ , bond angles and dihedral angles for protein side chains, and all degrees of freedom for the ligand. For the second stage, occupancies and estimates of binding free energies for the hydration sites are obtained using an additional 10–100 million MC configurations depending on the size and nature of the binding site.

In the case of neuraminidase below, the predictions of JAWS have been compared to those obtained by double-decoupling free energy simulations. For each water molecule to be annihilated, two free energy simulations are conducted. First the partial charges are turned off gradually and subsequently the LJ terms on the oxygen atom are perturbed to zero. The binding free energy is the sum of these two free energy changes minus the free energy of hydration of a TIP4P water molecule ( $-6.4$  kcal/mol). In all cases, evaluation of the potential energy employed 10- $\text{\AA}$  residue-based cutoffs. As mentioned above, the decoupled water molecule must be maintained in a well defined volume to guarantee convergence. For this purpose, a spherical hard-wall constraint of 2.8- $\text{\AA}$  radius was applied. The constraint was positioned at the beginning of the simulation at the center of the hydration site. The chosen water molecule was forbidden to escape its sphere of volume  $91.9 \text{\AA}^3$ . Furthermore, other water molecules were not permitted to diffuse into this excluded region, though solute and protein atoms could

occupy the space. The free energy calculations were carried out using 11 windows of simple overlap sampling.<sup>60,61</sup> Initial coordinates came from the preceding JAWS run. For each window, further equilibration began with 6 million MC configurations of water-only sampling. Full equilibration was then conducted for 4 million configurations, and averaging covered an additional 15 million configurations in each window.

## Results and Discussion

### Neuraminidase

The binding site of neuraminidase in complex with zanamivir (Figure 2) was selected as a first test case.<sup>62</sup> X-ray crystallography (1.80 Å) reveals that the ligand occupies only the top of the illustrated binding site, and it is accompanied by a well-resolved cluster of 6 water molecules. The binding site is isolated from the bulk solvent, so water molecules are not expected to enter or exit the binding site over the course of the simulations. Thus, this system provides an unambiguous test of the ability of JAWS to obtain the correct number of binding-site water molecules.

In fact, the algorithm identified 7 possible hydration sites, as shown in Figure 2A. Six of these sites match closely the positions of the crystallographic water molecules. Figure 2B shows the convergence of the fractional occupancy of the different hydration sites identified by clustering during phase 1 of the algorithm. The occupancy of 5 sites (sites 1, 3, 4, 5, 6) rapidly converged to values close to unity, while it took 4–5 million MC configurations to obtain stable occupancy values for sites 2 and 7. No other sites were identified in longer simulations of up to 20 million configurations. In Figure 2C, convergence plots for the free energies of binding of the water molecules at sites 4 and 7 are provided. Precise estimates for sites 1, 2, 3, 5, 6 can not be obtained after 50 million configurations because the  $\theta$ -water molecules located at these sites never sample  $\theta$  values below 0.05. However, the clear conclusion is that these sites should be fully occupied. For site 4, the  $\theta$ -water molecule turned off briefly, which enabled collection of sufficient statistics to allow estimation of the binding free energy. As illustrated, the binding free energy decreases slowly over the entire simulation, and site 4 is also computed to be occupied. The binding free energy estimate for the  $\theta$ -water molecule at site 7 oscillates around +2.5 kcal/mol, indicating that this site is not occupied. Thus, the hydration sites predicted by the JAWS method are fully consistent with the crystallographic data for this system.

As a key technical point, Figure 3 addresses the convergence of the fractional water occupancies as a function of the number of  $\theta$ -water molecules initially placed on the grid. For comparison, the results in Figure 2 were generated using 41  $\theta$ -water molecules. It is apparent that convergence is slower if the number of  $\theta$ -water molecules is lowered. It is important more  $\theta$ -water molecules are used than there are ultimate hydration sites (Figure 3C), otherwise some sites will go undetected. As the added computational cost of using higher numbers of  $\theta$ -water molecules is small, it is advisable to use an excess of  $\theta$ -water molecules. As a second item, the impact of the threshold at which a  $\theta$ -water molecule is considered “water-like” or “ideal-like” on the simulation results was assessed by changing the default from 0.95 to 0.90 or 0.98. The results were remarkably similar in all instances. In general, the  $\theta$ -water molecules are either completely “water-like” or “ideal-like” during most of the simulations and intermediate  $\theta$ -values are uncommon. An average  $\theta > 0.95$  is taken to indicate full occupancy of a site.

For comparison, Figure 4 shows the change in free energy of the system computed by double decoupling as sites 1–7 are incrementally filled with water molecules. The binding free energy decreases as each site is filled, before increasing again upon addition of a water molecule at site 7. The optimal number of water molecules is therefore 6 at sites 1 to 6, as in the crystal structure.<sup>62</sup> Occupancy of site 7 is predicted to be unfavorable by both JAWS and the double-decoupling calculations. From Figure 4, it is apparent that hydration of sites 4 and 5 is especially



favorable. Barillari et al. computed the absolute binding free energy of a water molecule at site 5 using a similar methodology.<sup>36</sup> While there are some differences between the two studies such as use of a different force field and protein setup, their reported binding free energy of  $-11.2 \pm 0.5$  kcal/mol is in good agreement with the present computed value ( $-11.4 \pm 0.2$  kcal/mol) using similar standard states.

Comparison of the JAWS and double decoupling results show that qualitatively both methods are in full agreement, predicting high water occupancy for sites 1 to 6 and no water occupancy of site 7. Quantitative comparison is imprecise since the methodologies do not compute binding affinities in the same fashion. Nevertheless as shown in Figure 4, the binding affinity of water for site 7 computed by double decoupling is ca. 5 kcal/mol, which compares with the ca. 3 kcal/mol computed by JAWS analysis (Figure 2C). For site 4, the water binding affinity by double decoupling is ca. -13 kcal/mol while the binding affinity predicted by JAWS analysis is continuously decreasing (ca. -4 kcal/mol after 50M). Comparisons for the other sites are not possible since the JAWS analysis did not provide statistics for low  $\theta$  values.

The aggregated CPU time for the double-decoupling FEP calculations was about 2300 hours on 3-GHz Pentium processors, while the JAWS analysis required about 15 hours, ca. 25% for phase 1 and 75% for phase 2. Furthermore, the setup of the double-decoupling calculations required *a priori* knowledge of the locations of the seven putative hydration sites, while these sites were automatically discovered by JAWS.

### Scytalone Dehydratase

Next, the JAWS method was applied to the complex of scytalone dehydratase with a cinnoline inhibitor (PDB code 3std, 1.65 Å).<sup>64</sup> This case was selected to test the performance on a binding site that is less polar than for neuraminidase. As summarized in Figure 5, 8 possible hydration sites were identified during phase 1; however, only 4 were computed to be significantly occupied by water molecules in phase 2. The same number is found in the crystal structure, and the computed and observed positions are all within 2 Å. Additionally, site 4 shows a binding affinity for water of ca. 0 kcal/mol and partial water occupancy of this site is therefore possible. Figure 6 shows the fluctuations in the value of  $\theta$  for hydration site 7, which is concluded to be unoccupied. This  $\theta$ -water molecule, as in the large majority of cases, rarely adopts intermediate  $\theta$  values that would correspond to an unphysical state that might contribute some inaccuracy to the free-energy estimates in phase 2.

### Major Urinary Protein 1

MUP1 has been the subject of intense scrutiny because ligand binding to this protein has been shown to involve an unusual enthalpy-driven hydrophobic effect.<sup>65</sup> It has been argued, partly based on MD simulations,<sup>66</sup> that this behavior is due to the poor hydration of the unliganded binding site. Hence, relatively few water molecules are expelled upon ligand binding and the binding affinity is dominated by ligand-protein interactions rather than entropy gain from the expulsion of ordered water molecules.<sup>67</sup>

A model of the binding site of MUP1 was setup from a single crystal structure of MUP1 at 1.60-Å resolution (PDB code 1znk).<sup>25</sup> JAWS analyses were then executed for two different ligands bound in this site with comparisons to the crystal structures for their MUP1 complexes. The initial configurations of the ligands were taken from their respective crystal structures. For the complex of MUP1 with 2-methoxy-3-isopropylpyrazine, the JAWS method yields 12 possible hydration sites (Figure 7A). Remarkably, none of these sites was found in phase 2 to have a negative binding free energy. This is consistent with the 1qy2 crystal structure (1.75 Å) for this ligand in which no ordered water molecules are observed.<sup>65</sup> However, when 2-*sec*-butyl-4,5-dihydrothiazole binds to MUP1, the JAWS analysis finds three hydration sites, 1, 2,

and 6, to be significantly occupied out of 8 possible sites (Figure 7B). Site 1 matches well the position of one site seen in the crystal structure (PDB code 1i06, 1.90 Å); this water molecule forms a hydrogen-bonding bridge between the hydroxyl group of Tyr120 and the backbone carbonyl of Leu40. Site 2 coincides with the position of a second water molecule seen in the crystal structure, which bridges between the first water molecule and the thiazolinylnitrogen of the ligand. However, in the simulations this water molecule is ca. 4.0 Å from the nitrogen atom of the thiazole ring and, hence, too far away for a strong hydrogen bond. Site 6 bridges between the hydroxyl group of Tyr120 and the thiazolinylnitrogen, but it was not reported in the crystal structure. In addition, there is one ambiguous site with a computed binding affinity ca. 0 kcal/mol, site 5, which can yield a hydrogen bond to the water molecule at site 6.

The discrepancy between the JAWS analyses and the experimental data may arise because, in starting from the 1znk structure, the thiazole ring is displaced to the right in Figure 7B (top) by about 1.5 Å in the simulations from its position observed in the 1i06 crystal structure. This reduces the importance of site 2 in favor of the site-6 water molecule that bridges between the thiazolinylnitrogen and Tyr120. To check this hypothesis, the JAWS simulations were repeated while constraining the ligand to maintain the 1i06 crystallographic binding mode. This resulted in full occupancy for site 1 and 2. All other hydration sites are predicted to be unfavorable. Thus, this case emerges as murky with potential partial water occupancies at sites 5 and 6.

Finally, JAWS analysis was used to derive the water content of the apo binding site. It was found that there are 18 possible hydration sites, but only 2 of them are predicted to be significantly occupied at room temperature, i.e., with affinities below -1.0 kcal/mol. Furthermore, many of the other sites had a binding affinity within 1 kcal/mol of zero suggesting that some of these sites could be transiently occupied. Overall, the results suggest that water molecules in the binding site of MUP1 are not well ordered and could be easily displaced by a ligand. The number of water molecules present ( $\theta > 0.95$ ) on average in the binding site during the simulations is 3.6. This matches well the estimate of 3.5 water molecules obtained by Barratt et al. from MD simulations.<sup>66</sup> Such a low number of water molecules leaves a substantial amount of space in the binding site, so the solvent density in the binding site is indeed low.

### Bovine $\beta$ -lactoglobulin

This protein binds large non-polar ligands in a deep hydrophobic cavity that is closed at low pH by a protonated Glu with an anomalously high  $pK_a$  (ca. 6.5). In basic media, the Glu becomes deprotonated and triggers a conformational change that opens up the lid of the binding site.<sup>68</sup> It has been proposed that, in the closed form of bovine  $\beta$ -lactoglobulin, a cluster of 5 water molecules occupies the ca. 320-Å<sup>3</sup> binding site, even in the absence of possible interactions with polar residues.<sup>68</sup> However, recent NMR and free-energy calculation studies have led to the conclusion that the binding site is not hydrated and that the electron density interpreted as 5 water molecules was probably due to a non-polar contaminant.<sup>69</sup>

A model of the binding site was constructed from a low-pH crystal structure (PDB code 3blg, 2.56 Å), and hydration sites were determined with the JAWS method. As reflected in Figure 8, 20 possible hydration sites were detected. However, the estimated binding affinities of water molecules at all of these sites are greater than +1 kcal/mol. Thus, the predictions from the JAWS analysis are in-line with the findings of Qvist et al., which also concluded that the binding site should not be hydrated.<sup>69</sup>

## Cyclooxygenase-2

COX-2 is another interesting case. Recent MD simulations have suggested that the binding site of apo COX-2 is dry, i.e., water molecules do not or only transiently occupy the Y-shaped channel.<sup>27</sup> While the binding site of bovine  $\beta$ -lactoglobulin appears to be dry, it is also exceptionally apolar. COX-2 represents a less extreme case, which includes some polar amino acid residues in the binding site, as indicated in Figure 9.<sup>69–70</sup>

Predictions for hydration sites were performed using the same structure for apo COX-2 as in the MD investigation of Young et al.;<sup>27</sup> the ligand, arachidonic acid, has been removed from the structure (PDB code 1cvu).<sup>70</sup> Twenty-seven possible hydration sites were identified during the first phase of the algorithm (Figure 9). Based on the number of grid points, the volume of the binding site is 400–450 Å<sup>3</sup>. As shown in Figure 10, most of the sites turn out to be uncompetitive with bulk hydration. No site is predicted to show a strong binding affinity for water at the end of the phase-2 simulations. There are, however, 16 sites with binding affinities within  $\pm 1$  kcal/mol of zero.

Similar results were obtained by repeating the JAWS calculations using alternative protocols. This suggests that the binding site of COX-2 is not completely free of water. The average number of water molecules present in the binding site is ca. 5. Together with the above estimate of the volume of the binding site this yields an average solvent density of about 0.3–0.4 g/ml in the binding site of COX-2. Owing to the large number of marginal sites, the results are sensitive to the choice of  $\Delta G_{hyd}(\text{water})$ ; a 0.5 kcal/mol lower value reduces the average number of water molecules in the binding site to only 1.2. The presence of a ligand can also affect the results. A previous MC study for COX2 complexes with celecoxib analogs indicated the common presence of bridging water molecules in the binding site, e.g., between Ser530 and Tyr385.<sup>71</sup>

As for MUP1, since there are multiple sites of low affinity, only a subset would be simultaneously occupied and the water molecules would mostly interact between themselves. Qualitatively, the results are clearly symptomatic of a binding site that is poorly hydrated by disordered, transient water molecules. The present results differ somewhat from the conclusion of Young et al. that the binding site is empty 80% of the time and that when seven water molecules were inserted in the binding site, they vacated it within 100 ps.<sup>27</sup> No experiments have provided further clarification. Given that the stabilities of several hydration sites are predicted to be marginal, small differences in the potential energy functions or simulation conditions may be responsible for some variation in the results.

## Conclusion

Application of the JAWS methodology produced reasonable predictions for the hydration of five small-molecule binding sites of varying polarity. In view of the use of the binding site grid, Monte Carlo sampling is the obvious choice for the JAWS simulations. Though the methodology is general, the results are expected to be sensitive to the choices for evaluating the molecular energetics. For example, since the non-polarizable TIP4P model of water was used in this work, some underestimation of the favorability of water molecules in a non-polar environment may be expected.<sup>69,72</sup> Compared with empirical techniques based on interaction energies or knowledge-based models such as GRID<sup>21</sup> or Superstar,<sup>24</sup> the JAWS approach has the advantage of explicitly incorporating entropic effects into the estimation of the binding free energies of water molecules. Nevertheless, the use of the  $\lambda$ -dynamics framework<sup>54</sup> and the umbrella sampling procedure<sup>55</sup> leads to efficient determination of the hydration pattern for binding sites, at a fraction of the computational cost required for a double-decoupling treatment.<sup>34</sup> The methods based on inhomogeneous fluid solvation theory also appear to be promising alternatives.<sup>26–28</sup>

For our environment, JAWS has been completely automated and produces a coordinate file that is suitable for further, standard MC, MD, and FEP calculations. However, attention to the setup of the grid is advisable. The JAWS methodology works particularly well for water molecules well-buried in cavities such that the grid is isolated from the bulk water. For binding sites that are solvent exposed, it may be preferable to forbid the diffusion of non  $\theta$ -water molecules into the grid in phase 2. It is reasonable, in view of the modest processing demands, to try more than one setup to seek consensus. For instance, in the case of MUP1, the putative hydration site 1 in Figure 7A is not seen in Figure 7B. For convenience, the grids were defined from the initial positions of the ligand atoms. The resulting grid covered the putative hydration site 1 with the pyrazine inhibitor, but not with the thiazole inhibitor. In this case, the difference was of no consequence because site 1 is not predicted to be hydrated. If for an application for which it is deemed important to use exactly the same grid for different ligands, it is easiest to define the grid using protein atoms. The results of a JAWS analysis also depend on the binding geometry of the ligand. If this information is not available from experimental data, JAWS can be used to derive the water content of the apo binding site. For docking calculations, water molecules predicted to be strongly bound can be retained, but allowed to be displaced by a ligand.<sup>14–17</sup>

Another important application of JAWS is to determine the water content of binding sites associated with computations of changes in free energies of protein-ligand binding and lead optimization. For a given ligand, the binding free energy estimates allow identification of water molecules whose replacement by ligand modifications could lead to significant activity gains.<sup>9</sup> JAWS results also allow addressing a perennial issue in FEP calculations on the presence or absence of specific water molecules in a binding site. It is desirable to perform a JAWS analysis on both endpoints prior to an FEP calculation. As an example, in an FEP study of celecoxib analogs binding to COX-2, the presence or absence of one water molecule was found to change the free-energy prediction for a phenyl-H to phenyl-F perturbation by 2.8 kcal/mol.<sup>71</sup> When the system setup was performed for the phenyl-H analog, the water molecule is introduced and becomes trapped during the H to F perturbation leading to overly disfavoring the fluoro derivative, while the water molecule is not present if the setup begins from the phenyl-F state. It is likely that the water molecule should not be present at all, which could be revealed by a JAWS analysis. Nonetheless, the results obtained with JAWS should be interpreted with care. In view of the use of the grid and the fact that the predicted binding affinities depend on the observation of a sufficient number of transitions between high and low  $\theta$  values, JAWS results are expected to be less precise than rigorous free energy calculations. While future implementation of adaptive biasing schemes to enhance convergence for the  $\theta$  parameters may be productive,<sup>73</sup> double-decoupling calculations remain advisable in ambiguous cases or for whenever refined quantitative results are desired. The utility of the JAWS approach is in rapidly locating putative hydration sites and in yielding semi-quantitative predictions about the affinity of water for such sites. Such knowledge is valuable in many contexts associated with the structure and activity of biomolecules and their complexes.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Gratitude is expressed to the National Institutes of Health (GM32136) and the National Foundation for Cancer Research for support of this research. J. M. also acknowledges support from a Marie Curie International Outgoing Fellowship from the European Commission (FP7-PEOPLE-2008-4-1-IOF, 234796-PPIdesign).

## References

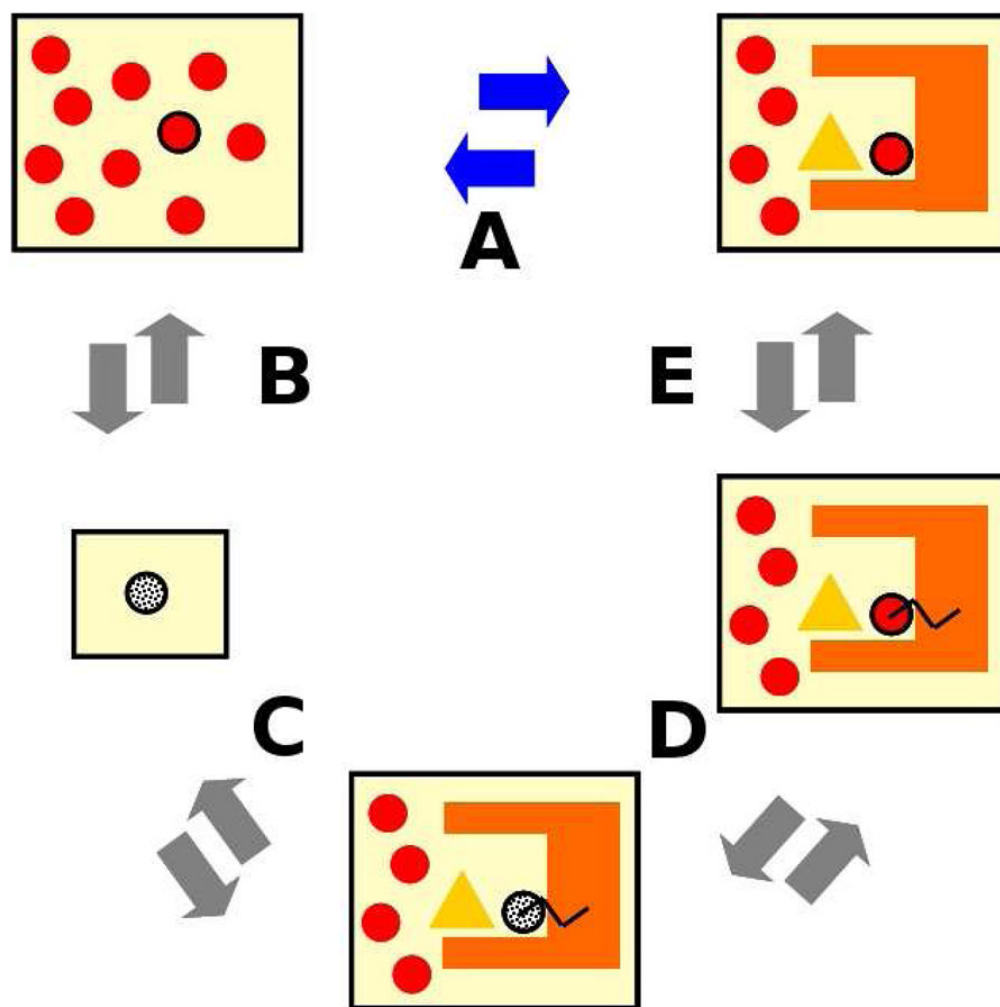
1. Gohlke H, Klebe G. *Angew Chem Int Ed Engl* 2002;41:2644–2676. [PubMed: 12203463]
2. (a) Jorgensen WL. *Science* 2004;303:1813–1818. [PubMed: 15031495] (b) Jorgensen WL. *Acc Chem Res* 2009;42:724–733. [PubMed: 19317443]
3. Chang CEA, Chen W, Gilson MK. *Proc Natl Acad Sci U S A* 2007;104:1534–1539. [PubMed: 17242351]
4. Tirado-Rives J, Jorgensen WL. *J Med Chem* 2006;49:5880–5884. [PubMed: 17004703]
5. Perola E, Charifson PS. *J Med Chem* 2004;47:2499–2510. [PubMed: 15115393]
6. Lazaridis T. *Curr Org Chem* 2002;6:1319–1332.
7. Southall NT, Dill KA, Haymet ADJ. *J Phys Chem B* 2002;106:521–533.
8. Lu YP, Wang RX, Yang CY, Wang SM. *J Chem Inf Model* 2007;47:668–675. [PubMed: 17266298]
9. (a) Lam PY, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang CH, Weber PC, Jackson DA, Sharpe TR, Erickson-Viitanen S. *Science* 1994;263:380–384. [PubMed: 8278812] (b) Ladbury JE. *Chem Biol* 1996;3:973–980. [PubMed: 9000013] (c) Li Z, Lazaridis T. *Phys Chem Chem Phys* 2007;9:573–581. [PubMed: 17242738] (d) Pan CF, Mezei M, Mujtaba S, Muller M, Zeng L, Li JM, Wang ZY, Zhou MM. *J Med Chem* 2007;50:2285–2288. [PubMed: 17444627]
10. Michel J, Tirado-Rives J, Jorgensen WL. 2009 submitted for publication.
11. Su Y, Gallicchio E, Das K, Arnold E, Levy RM. *J Chem Theory Comput* 2007;3:256–277.
12. (a) Michel J, Verdonk ML, Essex JW. *J Med Chem* 2006;49:7427–7439. [PubMed: 17149872] (b) Michel J, Taylor RD, Essex JW. *J Chem Theory Comput* 2006;2:732–739.
13. Michel J, Essex JW. *J Med Chem* 2008;51:6654–6664. [PubMed: 18834104]
14. Robeits BC, Mancera RL. *J Chem Inf Model* 2008;48:397–408. [PubMed: 18211049]
15. Huang N, Shoichet BK. *J Med Chem* 2008;51:4862–4865. [PubMed: 18680357]
16. Verdonk ML, Chessari G, Cole JC, Hartshorn MJ, Murray CW, Nissink JWM, Taylor RD, Taylor R. *J Med Chem* 2005;48:6504–6515. [PubMed: 16190776]
17. Rarey M, Kramer B, Lengauer T. *Proteins: Struct, Funct, Genet* 1999;34:17–28. [PubMed: 10336380]
18. Brown EN. *J Appl Crystallogr* 2008;41:761–767.
19. Carugo O, Bordo D. *Acta Crystallogr D Biol Crystallogr* 1999;55:479–483. [PubMed: 10089359]
20. Davis AM, St-Gallay SA, Kleywegt GJ. *Drug Discov Today* 2008;13:831–841. [PubMed: 18617015]
21. Goodford PJ. *J Med Chem* 1985;28:849–857. [PubMed: 3892003]
22. Gelpi JL, Kalko SG, Barril X, Cirera J, de la Cruz X, Luque FJ, Orozco M. *Proteins: Struct, Funct, Genet* 2001;45:428–437. [PubMed: 11746690]
23. Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. *J Mol Biol* 1997;265:445–464. [PubMed: 9034363]
24. Verdonk ML, Cole JC, Taylor R. *J Mol Biol* 1999;289:1093–1108. [PubMed: 10369784]
25. Malham R, Johnstone S, Bingham RJ, Barratt E, Phillips SEV, Laughton CA, Homans SW. *J Am Chem Soc* 2005;127:17061–17067. [PubMed: 16316253]
26. (a) Lazaridis T. *J Phys Chem* 1998;102:3531–3541. (b) Lazaridis T. *J Phys Chem* 1998;102:3542–3550. (c) Li Z, Lazaridis T. *J Phys Chem B* 2005;109:662–670. [PubMed: 16851059] (c) Li Z, Lazaridis T. *J Phys Chem B* 2006;110:1464–1475. [PubMed: 16471698]
27. Young T, Abel R, Kim B, Berne BJ, Friesner RA. *Proc Natl Acad Sci U S A* 2007;104:808–813. [PubMed: 17204562]
28. Abel R, Young T, Farid R, Berne BJ, Friesner RA. *J Am Chem Soc* 2008;130:2817–2831. [PubMed: 18266362]
29. (a) Resat H, Mezei M. *J Am Chem Soc* 1994;116:7451–7452. (b) Lynch GC, Pettitt BM. *J Chem Phys* 1997;107:8594–8610.
30. Mezei M. *Mol Phys* 1980;40:901–906.
31. Siepmann JI, Frenkel D. *Mol Phys* 1992;75:59–70.
32. Woo HJ, Dinner AR, Roux B. *J Chem Phys* 2004;121:6392–6400. [PubMed: 15446937]
33. Speidel JA, Banfelder JR, Mezei M. *J Chem Theory Comput* 2006;2:1429–1434.



34. (a) Wade RC, Mazor MH, McCammon JA, Quirocho FA. *J Am Chem Soc* 1990;112:7057–7059. (b) Gilson MK, Given JA, Bush BL, McCammon JA. *Biophys J* 1997;72:1047–1069. [PubMed: 9138555] (c) Hamelberg D, McCammon JA. *J Am Chem Soc* 2004;126:7683–7689. [PubMed: 15198616]
35. Zhang L, Hermans J. *Proteins: Struct Func Genet* 1996;24:433–438.
36. Barillari C, Taylor J, Viner R, Essex JW. *J Am Chem Soc* 2007;129:2577–2587. [PubMed: 17288418]
37. Barreiro G, Guimaraes CRW, Tubert-Brohman I, Lyons TM, Tirado-Rives J, Jorgensen WL. *J Chem Inf Model* 2007;47:2416–2428. [PubMed: 17949071]
38. Barreiro G, Kim JT, Guimaraes CRW, Bailey CM, Domaoal RA, Wang L, Anderson KS, Jorgensen WL. *J Med Chem* 2007;50:5324–5329. [PubMed: 17918923]
39. Jorgensen WL, Ruiz-Caro J, Tirado-Rives J, Basavapathruni A, Anderson KS, Hamilton AD. *Bioorg Med Chem Lett* 2006;16:663–667. [PubMed: 16263277]
40. Kim JT, Hamilton AD, Bailey CM, Domaoal RA, Wang LG, Anderson KS, Jorgensen WL. *J Am Chem Soc* 2007;129:3027–3027.
41. Ruiz-Caro J, Basavapathruni A, Kim JT, Bailey CM, Wang LG, Anderson KS, Hamilton AD, Jorgensen WL. *Bioorg Med Chem Lett* 2006;16:668–671. [PubMed: 16298131]
42. Thakur VV, Kim JT, Hamilton AD, Bailey CM, Domaoal RA, Wang LG, Anderson KS, Jorgensen WL. *Bioorg Med Chem Lett* 2006;16:5664–5667. [PubMed: 16931015]
43. Zeevaert JG, Wang LG, Thakur VV, Leung CS, Tirado-Rives J, Bailey CM, Domaoal RA, Anderson KS, Jorgensen WL. *J Am Chem Soc* 2008;130:9492–9499. [PubMed: 18588301]
44. Michel J, Harker EA, Tirado-Rives J, Jorgensen WL, Schepartz A. *J Am Chem Soc* 2009;131:6356–6357. [PubMed: 19415930]
45. Zwanzig RW. *J Chem Phys* 1954;22:1420–1426.
46. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J. *J Chem Phys* 1988;89:3742–3746.
47. Ben Naim, A. *Molecular Theory of Solutions*. Oxford University press; Oxford: 2006.
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys* 1983;79:926–935.
49. Jorgensen WL, Blake JF, Buckner JK. *Chem Phys* 1989;129:193–200.
50. Ben Naim A, Marcus Y. *J Chem Phys* 1984;81:2016–2027.
51. McQuarrie, DA. *Statistical Mechanics*. Harper & Row; New York: 1976.
52. Boresch S, Tettinger F, Leitgeb M, Karplus M. *J Phys Chem B* 2003;107:9535–9551.
53. Michel J, Verdonk ML, Essex JW. *J Chem Theory Comput* 2007;3:1645–1655.
54. Kong XJ, Brooks CL. *J Chem Phys* 1996;105:2414–2423.
55. Torrie GM, Valleau JP. *J Comput Phys* 1977;23:187–199.
56. Jorgensen WL, Tirado-Rives J. *J Comput Chem* 2005;26:1689–1700. [PubMed: 16200637]
57. (a) Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc* 1996;118:11225–11236. (b) Jorgensen WL, Tirado-Rives J. *Proc Natl Acad Sci U S A* 2005;102:6665–6670. [PubMed: 15870211]
58. Udier-Blagovic M, De Tirado PM, Pearlman SA, Jorgensen WL. *J Comput Chem* 2004;25:1322–1332. [PubMed: 15185325]
59. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. *J Chem Phys* 1953;21:1087–1092.
60. Jorgensen WL, Thomas LL. *J Chem Theory Comput* 2008;4:869–876. [PubMed: 19936324]
61. Lu ND, Kofke DA, Woolf TB. *J Comput Chem* 2004;25:28–39. [PubMed: 14634991]
62. Varghese JN, Epa VC, Colman PM. *Protein Sci* 1995;4:1081–1087. [PubMed: 7549872]
63. Humphrey W, Dalke A, Schulten K. *J Mol Graph* 1996;14:33–38. [PubMed: 8744570]
64. Chen JM, Xu SL, Wawrzak Z, Basarab GS, Jordan DB. *Biochemistry* 1998;37:17735–17744. [PubMed: 9922139]
65. Bingham RJ, Findlay JBC, Hsieh SY, Kalverda AP, Kjeliberg A, Perazzolo C, Phillips SEV, Seshadri K, Trinh CH, Turnbull WB, Bodenhausen G, Homans SW. *J Am Chem Soc* 2004;126:1675–1681. [PubMed: 14871097]
66. Barratt E, Bingham RJ, Warner DJ, Laughton CA, Phillips SEV, Homans SW. *J Am Chem Soc* 2005;127:11827–11834. [PubMed: 16104761]

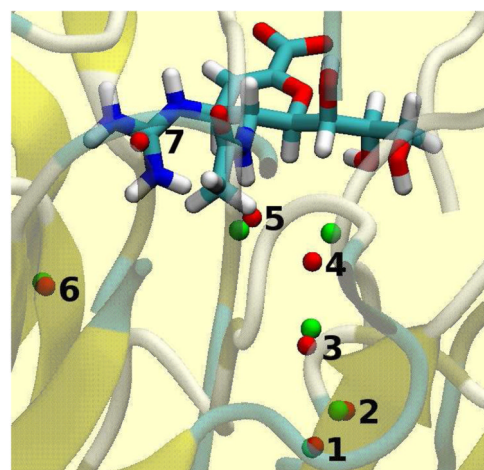


67. Homans SW. *Drug Discov Today* 2007;12:534–539. [PubMed: 17631247]
68. Qin BY, Bewley MC, Creamer LK, Baker HM, Baker EN, Jameson GB. *Biochemistry* 1998;37:14014–14023. [PubMed: 9760236]
69. Qvist J, Davidovic M, Hamelberg D, Halle B. *Proc Natl Acad Sci U S A* 2008;105:6296–6301. [PubMed: 18427121]
70. Kiefer JR, Pawlitz JL, Moreland KT, Stegeman RA, Hood WF, Gierse JK, Stevens AM, Goodwin DC, Rowlinson SW, Marnett LJ, Stallings WC, Kurumbail RG. *Nature* 2000;405:97–101. [PubMed: 10811226]
71. Price MLP, Jorgensen WL. *J Am Chem Soc* 2000;122:9455–9466.
72. Jorgensen WL, McDonald NA, Selmi M, Rablen PR. *J Am Chem Soc* 1995;117:11809–11810.
73. Mezei M. *J Comput Phys* 1987;68:237–248.

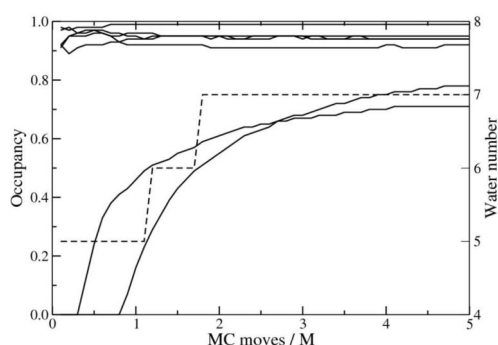


**Figure 1.**

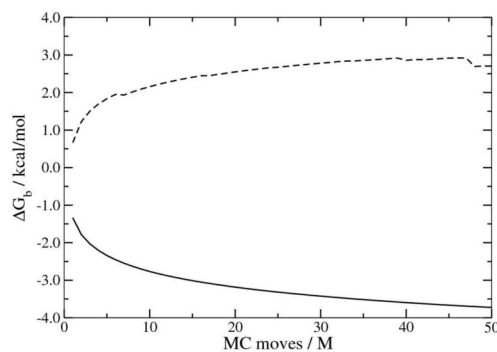
Reversible transfer of a water molecule from bulk solvent to a well defined position in a protein binding site. A) Direct process. B) Decoupling of one water molecule from bulk. C) Localization of an ideal particle into a binding site. D) Conversion of a localized ideal particle into a water molecule. E) Removal of the constraint. Water molecules are depicted by red spheres, ideal particles by gray spheres. The orange shape represents a binding site. The yellow triangle represents a ligand. The black zigzag depicts a volume constraint.



(A)



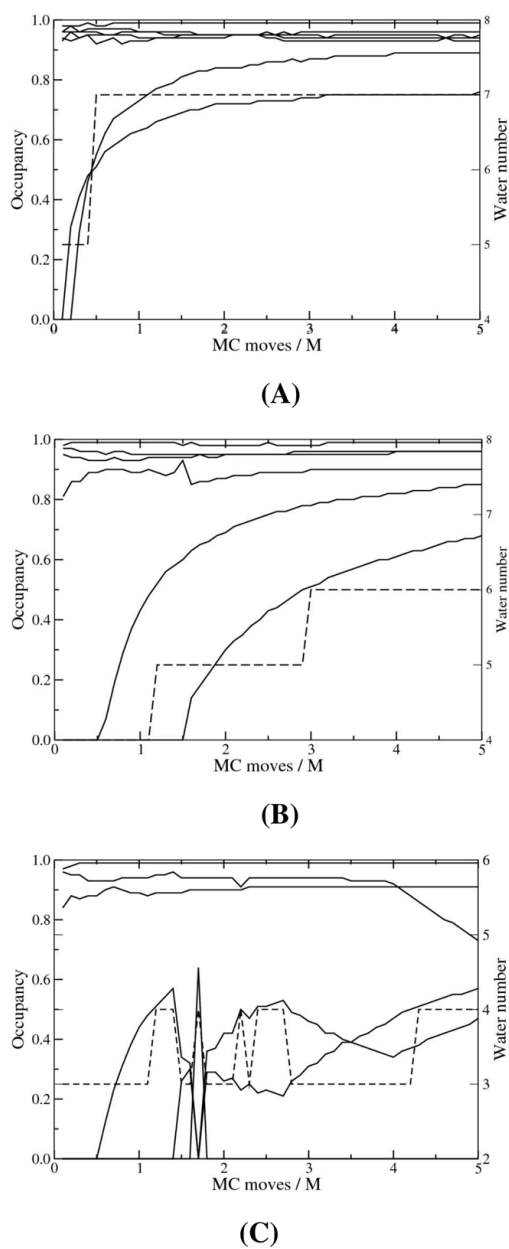
(B)



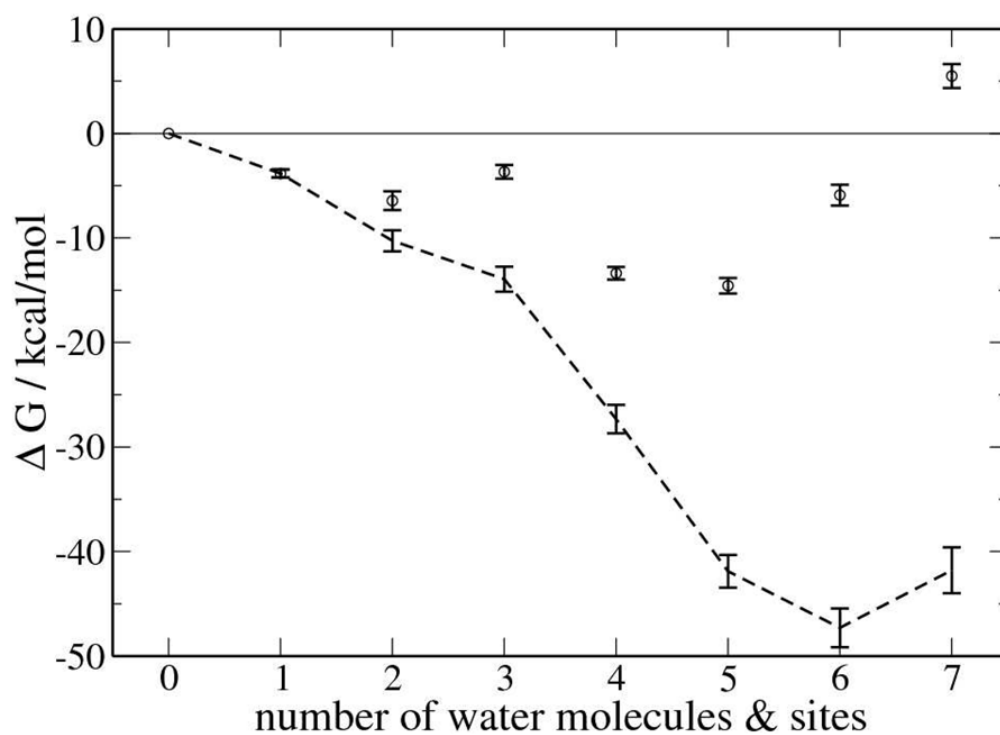
(C)

**Figure 2.**

Detection of hydration sites in the binding site of neuraminidase complexed with zanamivir. (A) In red, hydration sites identified by JAWS. In green, location of ordered water molecules observed in the crystal structure (PDB code 1nnc).<sup>62</sup> Zanamivir is shown in sticks, the protein binding site is shown in cartoon representation. The figure was created with VMD.<sup>63</sup> (B) Occupancy probabilities of the possible hydration sites discovered in the binding site of neuraminidase. The dashed line shows the number of sites whose occupancy probability is greater than 0.5. (C) Binding free energy estimates for sites 4 (black line) and 7 (dashed line).

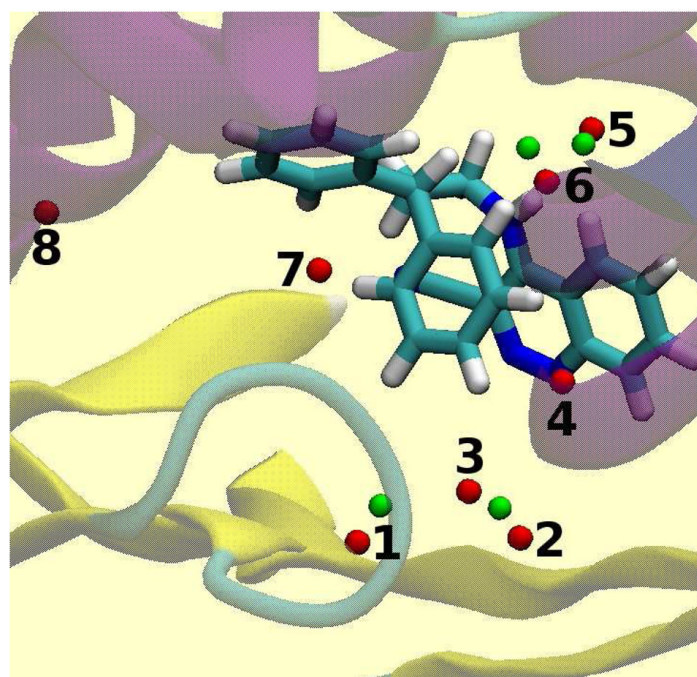


**Figure 3.** Occupancy probability of the possible hydration sites in the binding site of the neuraminidase-zanamivir complex using different numbers of  $\theta$ -water molecules: (A) 20, (B) 10, and (C) 5  $\theta$ -water molecules. The dashed line records the number of sites with occupancy probability greater than 0.5.

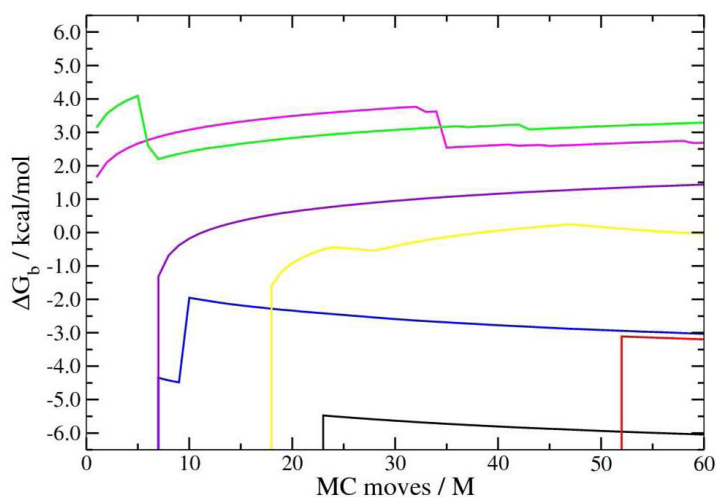


**Figure 4.**

Free energy changes from double-decoupling for the addition of one to seven water molecules in the binding site of the neuraminidase-zanamivir complex. The x-axis shows the number and position of the water molecules, e.g., 3 means three water molecules at sites 1, 2, and 3. Circles denote the free energy change upon addition of each water molecule and the dashed line shows the cumulative free energy change. For clarity, the error bars have been expanded; they represent  $\pm 5$  standard deviations.



(A)

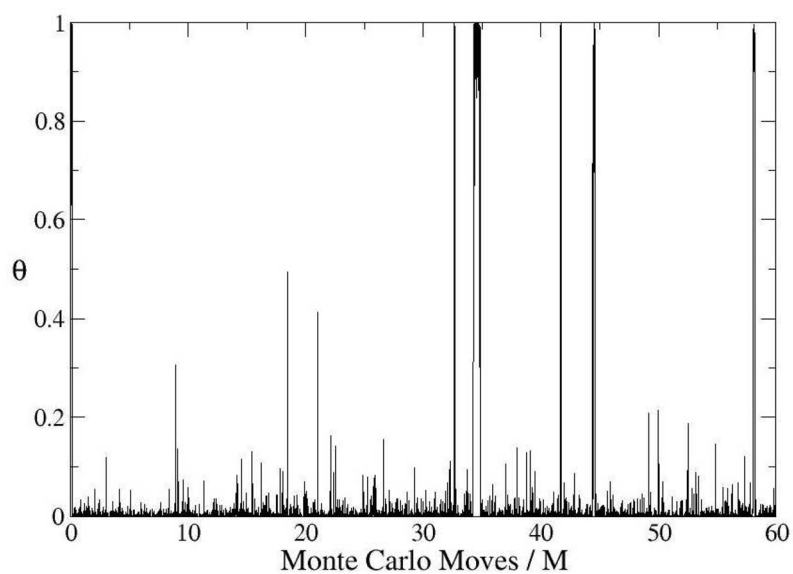


(B)

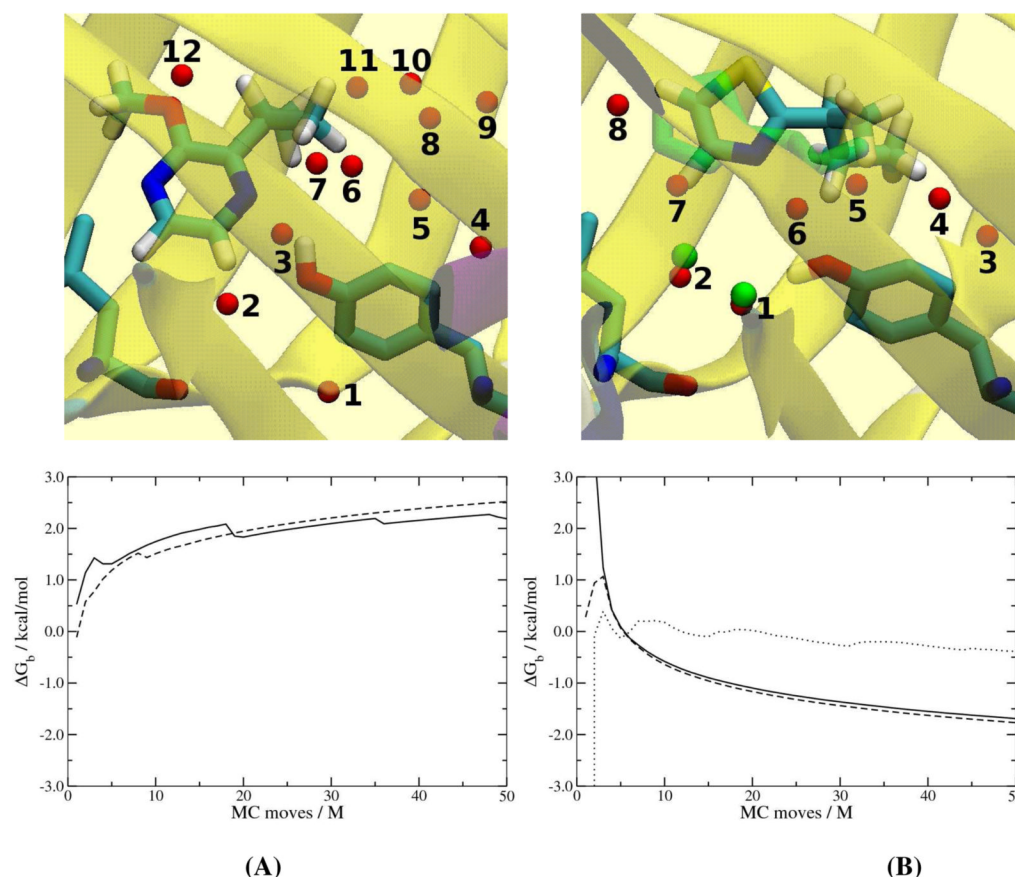
**Figure 5.**

Hydration sites in the binding site of scytalone dehydratase complexed to a cinnoline inhibitor. (A) In red, the possible hydration sites identified by JAWS. In green, the positions of ordered water molecules in the 3std crystal structure. (B) Binding free energy estimates for the putative hydration sites. Site 1 (black), site 3 (green), site 4 (yellow), site 5 (red), site 6 (blue), site 7 (magenta), site 8 (violet). Since the  $\theta$ -water molecule at site 2 never turned off, its binding free energy was not determined.

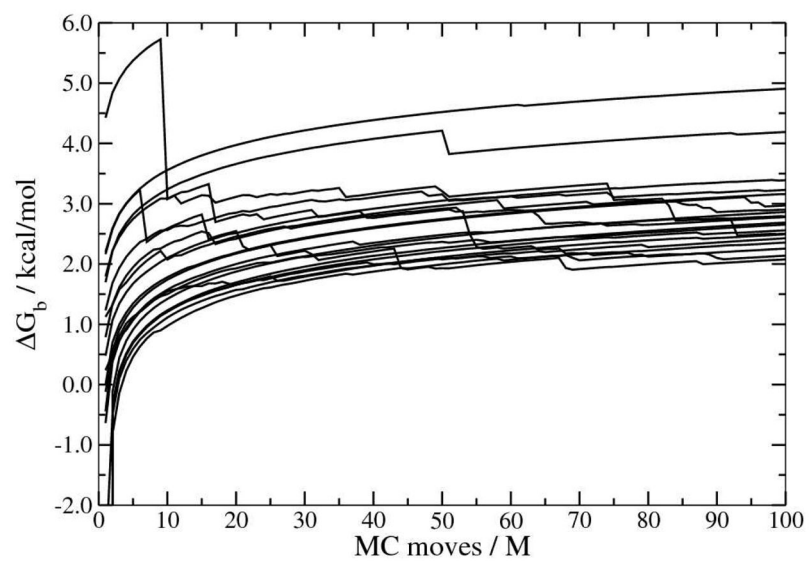




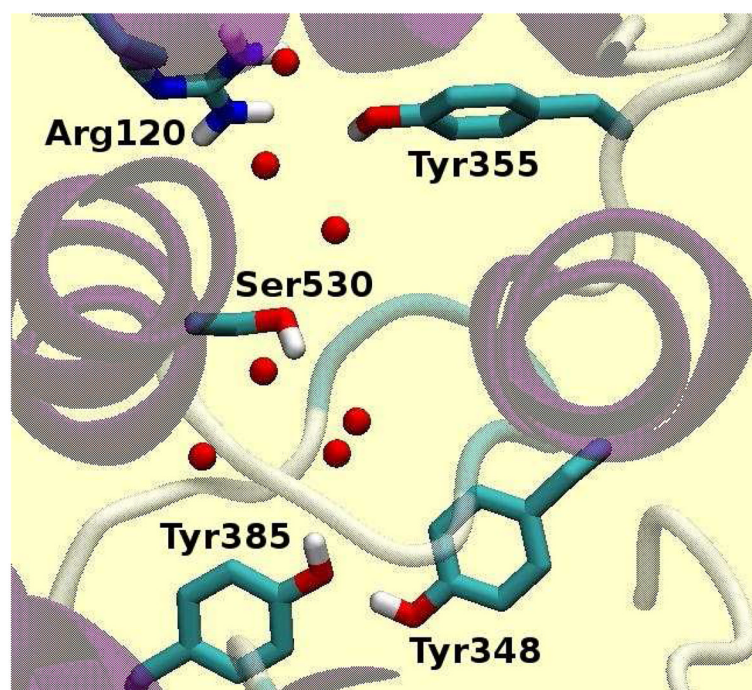
**Figure 6.** Fluctuations in the value of  $\theta$  for the  $\theta$ -water molecule at site 7 in the binding site of the scytalone dehydratase complex.



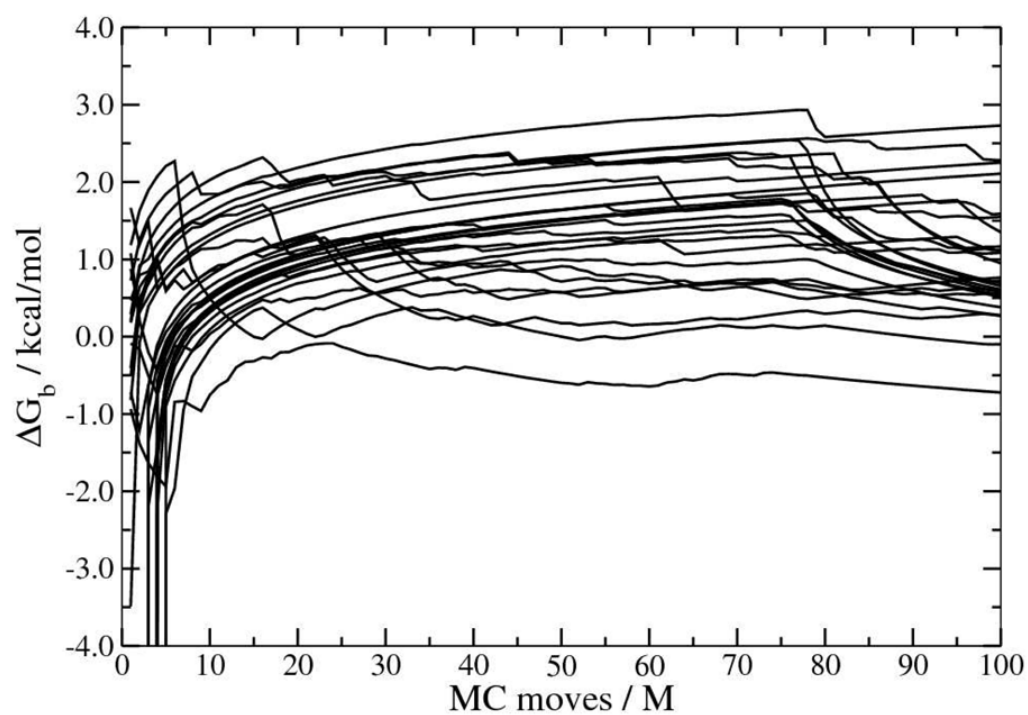
**Figure 7.** Hydration sites in the binding site of MUP1 complexes. (A) Predictions with the ligand 2-methoxy-3-isopropylpyrazine. (B) Predictions with the ligand 2-*sec*-butyl-4,5-dihydrothiazole. The location of the ligand in the crystal structure is shown in green. Top: In red, the possible hydration sites identified by JAWS. In green, the position of ordered water molecules observed in the crystal structures. The ligand and residues Leu40 and Tyr120 are represented in sticks. Bottom: Binding affinity estimates for selected sites. (A) Sites 2 (black line) and 3 (dashed line). (B) Sites 1 (black line), 2 (dashed line), 5 (dotted line). Since the  $\theta$ -water at site 6 never turned off its binding free energy was not determined.



**Figure 8.** Binding free energy estimates for 20 possible hydration sites identified in the binding site of bovine  $\beta$ -lactoglobulin.



**Figure 9.** Some of the polar residues in the COX-2 binding site; Arg120 and Tyr355 are at the entrance to the site. The 7 most likely hydration sites found by the JAWS analysis are indicated with red balls; however, the computed binding energies are weak, ranging from  $-0.7$  to  $0.5$  kcal/mol.



**Figure 10.**  
Estimated binding free energies for 27 possible hydration sites identified in the binding site of COX-2.